

**UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF TEXAS  
SHERMAN DIVISION**

R2 SOLUTIONS LLC,  
*Plaintiff*

v.

DATABRICKS, INC.,  
*Defendant*

§  
§  
§  
§  
§  
§  
§

CIVIL ACTION NO. 4:23-CV-1147  
(Judge Mazzant)

**CLAIM CONSTRUCTION MEMORANDUM OPINION AND ORDER**

Before the Court is Plaintiff R2 Solutions LLC’s (“Plaintiff’s” or “R2’s”) Opening Claim Construction Brief (Dkt. #53). Also before the Court is the Responsive Claim Construction Brief (Dkt. #62) filed by Defendant (“Defendant” or “Databricks”), as well as Plaintiff’s reply (Dkt. #66). Further before the Court are the parties’ September 13, 2024 P.R. 4-3 Joint Claim Construction Statement (Dkt. #51) and the parties’ November 22, 2024 P.R. 4-5(d) Joint Claim Construction Chart (Dkt. #69, Ex. A).

The Court held a claim construction hearing on December 2, 2024, to determine the proper construction of the disputed claim terms in United States Patent No. 8,190,610.

The Court issues this Claim Construction Memorandum Opinion and Order and hereby incorporates-by-reference the claim construction hearing transcript.

**BACKGROUND**

Plaintiff alleges infringement of United States Patent No. 8,190,610 (“the ’610 Patent” or “the patent-in-suit”). The ’610 Patent relates to computer data processing. The Abstract of the ’610 Patent states:

An input data set is treated as a plurality of grouped sets of key/value pairs, which enhances the utility of the MapReduce programming methodology. By utilizing

such a grouping, map processing can be carried out independently on two or more related but possibly heterogeneous datasets (e.g., related by being characterized by a common primary key). The intermediate results of the map processing (key/value pairs) for a particular key can be processed together in a single reduce function by applying a different iterator to intermediate values for each group. Different iterators can be arranged inside reduce functions in ways however desired.

The Court previously construed disputed terms in the '610 Patent in:

*R2 Solutions LLC v. Deezer, S.A., et al.*, Nos. 4:21-CV-90, 4:21-CV-91, 4:21-CV-122, 4:21-CV-174, Dkt. #54 (E.D. Tex. Jan. 4, 2022) (referred to as “Deezer” or as “Walmart”); and

*R2 Solutions LLC v. American Airlines, Inc.*, No. 4:22-CV-353-ALM, Dkt. #58 (E.D. Tex. Apr. 3, 2023) (referred to as “American Airlines”).

### LEGAL STANDARDS

Claim construction is a matter of law. *Markman v. Westview Instruments, Inc.*, 52 F.3d 967, 979 (Fed. Cir. 1995). The purpose of claim construction is to resolve the meanings and technical scope of claim terms. *U.S. Surgical Corp. v. Ethicon, Inc.*, 103 F.3d 1554, 1568 (Fed. Cir. 1997). When the parties dispute the scope of a claim term, “it is the court’s duty to resolve it.” *O2 Micro Int’l Ltd. v. Beyond Innovation Tech. Co.*, 521 F.3d 1351, 1362 (Fed. Cir. 2008).

“It is a ‘bedrock principle’ of patent law that ‘the claims of a patent define the invention to which the patentee is entitled the right to exclude.’” *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005) (quoting *Innova/Pure Water, Inc. v. Safari Water Filtration Sys., Inc.*, 381 F.3d 1111, 1115 (Fed. Cir. 2004)). The Court examines a patent’s intrinsic evidence to define the patented invention’s scope. *Id.* at 1313–14; *Bell Atl. Network Servs., Inc. v. Covad Commc’ns Group, Inc.*, 262 F.3d 1258, 1267 (Fed. Cir. 2001). Intrinsic evidence includes the claims, the rest of the specification, and the prosecution history. *Phillips*, 415 F.3d at 1312–13; *Bell Atl. Network Servs.*, 262 F.3d at 1267. The Court gives claim terms their ordinary and customary meaning as

understood by one of ordinary skill in the art at the time of the invention. *Phillips*, 415 F.3d at 1312–13; *Alloc, Inc. v. Int’l Trade Comm’n*, 342 F.3d 1361, 1368 (Fed. Cir. 2003).

Claim language guides the Court’s construction of claim terms. *Phillips*, 415 F.3d at 1314. “[T]he context in which a term is used in the asserted claim can be highly instructive.” *Id.* Other claims, asserted and unasserted, can provide additional instruction because “terms are normally used consistently throughout the patent.” *Id.* Differences among claims, such as additional limitations in dependent claims, can provide further guidance. *Id.*

“[C]laims ‘must be read in view of the specification, of which they are a part.’” *Id.* at 1315 (quoting *Markman*, 52 F.3d at 979). “[T]he specification ‘is always highly relevant to the claim construction analysis. Usually, it is dispositive; it is the single best guide to the meaning of a disputed term.’” *Id.* (quoting *Vitronics Corp. v. Conceptronic, Inc.*, 90 F.3d 1576, 1582 (Fed. Cir. 1996)); *Teleflex, Inc. v. Ficoso N. Am. Corp.*, 299 F.3d 1313, 1325 (Fed. Cir. 2002). In the specification, a patentee may define his own terms, give a claim term a different meaning than it would otherwise possess, or disclaim or disavow some claim scope. *Phillips*, 415 F.3d at 1316. Although the Court generally presumes terms possess their ordinary meaning, this presumption can be overcome by statements of clear disclaimer. *See SciMed Life Sys., Inc. v. Advanced Cardiovascular Sys., Inc.*, 242 F.3d 1337, 1343–44 (Fed. Cir. 2001). This presumption does not arise when the patentee acts as his own lexicographer. *See Irdeto Access, Inc. v. EchoStar Satellite Corp.*, 383 F.3d 1295, 1301 (Fed. Cir. 2004).

The specification may also resolve ambiguous claim terms “where the ordinary and accustomed meaning of the words used in the claims lack sufficient clarity to permit the scope of the claim to be ascertained from the words alone.” *Teleflex*, 299 F.3d at 1325. For example, “[a] claim interpretation that excludes a preferred embodiment from the scope of the claim ‘is rarely,

if ever, correct.” *Globetrotter Software, Inc. v. Elan Computer Group Inc.*, 362 F.3d 1367, 1381 (Fed. Cir. 2004) (quoting *Vitronics*, 90 F.3d at 1583). But, “[a]lthough the specification may aid the court in interpreting the meaning of disputed language in the claims, particular embodiments and examples appearing in the specification will not generally be read into the claims.” *Constant v. Advanced Micro-Devices, Inc.*, 848 F.2d 1560, 1571 (Fed. Cir. 1988); *accord Phillips*, 415 F.3d at 1323.

The prosecution history is another tool to supply the proper context for claim construction because a patentee may define a term during prosecution of the patent. *Home Diagnostics Inc. v. LifeScan, Inc.*, 381 F.3d 1352, 1356 (Fed. Cir. 2004) (“As in the case of the specification, a patent applicant may define a term in prosecuting a patent.”). The well-established doctrine of prosecution disclaimer “preclud[es] patentees from recapturing through claim interpretation specific meanings disclaimed during prosecution.” *Omega Eng’g Inc. v. Raytek Corp.*, 334 F.3d 1314, 1323 (Fed. Cir. 2003). “Indeed, by distinguishing the claimed invention over the prior art, an applicant is indicating what the claims do not cover.” *Spectrum Int’l v. Sterilite Corp.*, 164 F.3d 1372, 1378–79 (Fed. Cir. 1988) (quotation omitted). “As a basic principle of claim interpretation, prosecution disclaimer promotes the public notice function of the intrinsic evidence and protects the public’s reliance on definitive statements made during prosecution.” *Omega Eng’g*, 334 F.3d at 1324. However, the prosecution history must show that the patentee clearly and unambiguously disclaimed or disavowed the proposed interpretation during prosecution to obtain claim allowance. *Middleton Inc. v. 3M Co.*, 311 F.3d 1384, 1388 (Fed. Cir. 2002). Statements will constitute disclaimer of scope only if they are “clear and unmistakable statements of disavowal.” *See Cordis Corp. v. Medtronic AVE, Inc.*, 339 F.3d 1352, 1358 (Fed. Cir. 2003). An “ambiguous disavowal”

will not suffice. *Schindler Elevator Corp. v. Otis Elevator Co.*, 593 F.3d 1275, 1285 (Fed. Cir. 2010) (citation omitted).

Although “less significant than the intrinsic record in determining the legally operative meaning of claim language,” the Court may rely on extrinsic evidence to “shed useful light on the relevant art.” *Phillips*, 415 F.3d at 1317 (quotation omitted). Technical dictionaries and treatises may help the Court understand the underlying technology and the manner in which one skilled in the art might use claim terms, but such sources may also provide overly broad definitions or may not be indicative of how terms are used in the patent. *Id.* at 1318. Similarly, expert testimony may aid the Court in determining the particular meaning of a term in the pertinent field, but “conclusory, unsupported assertions by experts as to the definition of a claim term are not useful.” *Id.* Generally, extrinsic evidence is “less reliable than the patent and its prosecution history in determining how to read claim terms.” *Id.*

The Supreme Court of the United States has “read [35 U.S.C.] § 112, ¶ 2 to require that a patent’s claims, viewed in light of the specification and prosecution history, inform those skilled in the art about the scope of the invention with reasonable certainty.” *Nautilus, Inc. v. Biosig Instruments, Inc.*, 134 S. Ct. 2120, 2129 (2014). “A determination of claim indefiniteness is a legal conclusion that is drawn from the court’s performance of its duty as the construer of patent claims.” *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342, 1347 (Fed. Cir. 2005) (citations and internal quotation marks omitted), *abrogated on other grounds by Nautilus*, 134 S. Ct. 2120. “Indefiniteness must be proven by clear and convincing evidence.” *Sonix Tech. Co. v. Publ’ns Int’l, Ltd.*, 844 F.3d 1370, 1377 (Fed. Cir. 2017).

## ANALYSIS

### *Agreed Claim Terms*

In their September 13, 2024 P.R. 4-3 Joint Claim Construction and Pre-Hearing Statement (Dkt. #51, at 1–2) and in their November 22, 2024 P.R. 4-5(d) Joint Claim Construction Chart (Dkt. #69, Ex. A, at pp. 9–11), the parties submitted the following agreed-upon constructions, which the Court hereby adopts as agreed-upon:

<u>Term</u>	<u>Agreed Construction</u>
“A method of processing data of a data set over a distributed system, wherein the data set comprises a plurality of data groups, the method comprising:” (Claim 1)	Limiting preamble in Claim 1.
“A computer system including a plurality of computing devices, the computer system configured to process data of a data set, wherein the data set comprises a plurality of data groups, the computer system comprises at least one processor and memory that are operable to perform the following operations:” (Claim 17)	Limiting preamble in Claim 17.
“data group” (Claims 1–5, 17–21)	“a group of data and a mechanism for identifying data from that group”
“plurality of mapping functions that are each user-configurable” (Claims 1, 17)	“two or more mapping functions that are each configurable by a user”

*Disputed Claim Terms*

1. “processor and memory that are operable to perform the following operations:  
 . . . based on the key in common”

<p><b>“processor and memory that are operable to perform the following operations:          . . . based on the key in common”          (Claim 17)</b></p>	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
<p>Plain and ordinary meaning. Not subject to 35 U.S.C. § 112, ¶ 6 and not indefinite.</p> <p>See also the proposed constructions for “data group” and “a plurality of mapping functions that are each user-configurable.”</p>	<p>Governed by pre-AIA 35 U.S.C. § 112, ¶ 6.</p> <p>Function:</p> <p>“partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of mapping functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such map function’s corresponding data partition to form corresponding intermediate data for that data group and identifiable to that data group, wherein the data of a first data group has a different schema than the data of a second data group and the data of the first data group is mapped differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different schema and corresponding different intermediate data have a key in common; and reduce the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group so as to result in a merging of the corresponding different intermediate data based on the key in common.”</p> <p>Structure: Indefinite.</p>

(Dkt. #51, Ex. A, at p. 1; *id.*, Ex. B, at pp. 1–3; Dkt. #53, at p. 5; Dkt. #62, at pp. 3–4; Dkt. #69, Ex. A, at pp. 1–2).

**a. The Parties' Positions**

Plaintiff argues that “Databricks fails to overcome the presumption that the phrase is not [] governed by 35 U.S.C. § 112, ¶ 6, and further fails to establish indefiniteness.” (Dkt. #53, at p. 5). Plaintiff argues that the claim itself outlines objectives and operations, and that the specification, as well as the prosecution history, demonstrate that processors and memory were well-known. (*Id.*, at pp. 6–10). Alternatively, Plaintiff submits that, if the Court finds that § 112, ¶ 6 applies, then “the Court should find the phrase not indefinite because structure is found in at least Figures 4 and 5 and the accompanying discussion.” (*Id.*, at p. 10).

Defendant responds that “‘processor and memory that are operable to perform the following operations’ fails to connote sufficient structure for performing the claimed function,” and “[t]he term is indefinite because the patent discloses no steps or algorithm clearly linked to the term.” (Dkt. #62, at p. 4; *see id.*, at pp. 13–17). Defendant urges that “the claim is purely functional and recites the *results* of ‘partitioning,’ ‘mapping,’ and ‘reducing’ without identifying sufficient structure for the entirety of the claimed function.” (*Id.*, at p. 9). Defendant also argues that “[b]ecause a generic processor and memory can only provide structure for their generic functions like accessing, storing, and retrieving data, they cannot provide structure for the specialized function here.” (*Id.*, at p. 11) (citation omitted).

Plaintiff replies that “Claim 17 recites the algorithm performed by the ‘processor and memory,’” and “a POSA does not need the patent to explain how processors and memory work and communicate.” (Dkt. #66, at p. 1 & 3).



## b. Analysis

Title 35 U.S.C. § 112(f) (formerly § 112, ¶ 6) provides: “An element in a claim for a combination may be expressed as a means or step for performing a specified function without the recital of structure, material, or acts in support thereof, and such claim shall be construed to cover the corresponding structure, material, or acts described in the specification and equivalents thereof.” “In exchange for using this form of claiming, the patent specification must disclose with sufficient particularity the corresponding structure for performing the claimed function and clearly link that structure to the function.” *Triton Tech of Tex., LLC v. Nintendo of Am., Inc.*, 753 F.3d 1375, 1378 (Fed. Cir. 2014).

“[T]he failure to use the word ‘means’ . . . creates a rebuttable presumption . . . that § 112, para. 6 does not apply.” *Williamson v. Citrix Online LLC*, 792 F.3d 1339, 1348 (Fed. Cir. 2015) (citations and internal quotation marks omitted). “When a claim term lacks the word ‘means,’ the presumption can be overcome and § 112, para. 6 will apply if the challenger demonstrates that the claim term fails to recite sufficiently definite structure or else recites function without reciting sufficient structure for performing that function.” *Id.* at 1349 (citations and internal quotation marks omitted).

*Williamson*, in an *en banc* portion of the decision, abrogated prior statements that the absence of the word “means” gives rise to a “strong” presumption against means-plus-function treatment. *Id.* (citation omitted). *Williamson* also abrogated prior statements that this presumption “is not readily overcome” and that this presumption cannot be overcome “without a showing that the limitation essentially is devoid of anything that can be construed as structure.” *Id.* (citations omitted). Instead, *Williamson* found, “[h]enceforth, we will apply the presumption as we have done prior to *Lighting World* . . . .” *Id.* (citing *Lighting World, Inc. v. Birchwood Lighting, Inc.*,

382 F.3d 1354, 1358 (Fed. Cir. 2004)). In a subsequent part of the decision not considered *en banc*, *Williamson* affirmed the district court’s finding that the term “distributed learning control module” was a means-plus-function term that was indefinite because of lack of corresponding structure, and in doing so *Williamson* stated that “‘module’ is a well-known nonce word.” 792 F.3d at 1350.

Here, Claim 17 of the ’610 Patent recites (emphasis added):

17. A computer system including a plurality of computing devices, the computer system configured to process data of a data set, wherein the data set comprises a plurality of data groups, the computer system comprises at least one *processor* and *memory* that are operable to perform the following operations:

partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of mapping functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such map function’s corresponding data partition to form corresponding intermediate data for that data group and identifiable to that data group, wherein the data of a first data group has a different schema than the data of a second data group and the data of the first data group is mapped differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different schema and corresponding different intermediate data have a key in common; and

reduce the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group so as to result in a merging of the corresponding different intermediate data based on the key in common.

This “processor and memory” term does not use any of the words identified by *Williamson* as a “nonce” word lacking structure. *See id.* Although “processor” and “memory” may refer to broad classes of structures, this breadth does not necessarily render the term non-structural. *See Skky, Inc. v. MindGeek, s.a.r.l.*, 859 F.3d 1014, 1019 (Fed. Cir. 2017) (finding “wireless device means” not a means-plus-function term, noting that “it is sufficient if the claim term is used in common parlance or by persons of skill in the pertinent art to designate structure, even if the term covers a broad class of structures and even if the term identifies the structures by their function”) (quoting *TecSec, Inc. v. Int’l Bus. Machs. Corp.*, 731 F.3d 1336, 1347 (Fed. Cir. 2013)).

Moreover, the entire body of the claim sets forth the objectives and operations of the processor. *See Linear Tech. Corp. v. Impala Linear Corp.*, 379 F.3d 1311, 1319–21 (Fed. Cir. 2004). Thus, the claim language itself provides sufficiently definite structure to one of skill in the art. *See id.* Defendant argues that *Linear Tech* is inapplicable because, in Defendant’s view, Plaintiff does not submit sufficient expert opinion. (Dkt. #63, at p. 10). Defendant’s argument is unpersuasive because the objectives and operations of the processor are recited in detail in the claim itself, as reproduced above.

Also, as Plaintiff submits, prior art cited during prosecution includes discussion of processors and memory as being structural. (*See* Dkt. #53, Ex. 2, Ex. G, U.S. Patent No. 6,341,289 at 6:43–51) (p. 582 of 611 of Ex. 2) (“Processor 10 may comprise a single integrated circuit, such as a microprocessor, or may comprise any suitable number of integrated circuit devices and/or circuit boards working in cooperation to accomplish the functions of a processor.”); *see also id.* at 7:26–41 (“[M]emory 120 is used in its broadest sense, and can include Dynamic Random Access Memory (DRAM), Static RAM (SRAM), flash memory, cache memory, etc.”; “memory 120 may be a single type of memory component or may be composed of many different types of memory components”; “memory 120 and CPU 110 may be distributed across several different computers that collectively comprise system 100”).

Defendant cites the Court’s decision in *St. Isidore*, which found that the presumption against means-plus-function treatment under 35 U.S.C. § 112, ¶ 6 had been rebutted because “[i]n the context of the ‘processor configured to . . .’ terms, . . . each processor is defined only by the function that it performs.” *St. Isidore Research, LLC v. Comerica Inc.*, No. 2:15-CV-1390-JRG-RSP, 2016 WL 4988246, at \*14 (E.D. Tex. Sept. 19, 2016) (Payne, J.). *St. Isidore* itself noted that “[t]he Court has typically found ‘processor’ to connote sufficient structure to avoid the application

of § 112, ¶ 6 in different circumstances.” *Id.*, at \*15. The Court’s analysis in *SyncPoint* is applicable. *See SyncPoint Imaging, LLC v. Nintendo of Am. Inc.*, No. 2:15-CV-247, 2016 WL 55118, at \*20 (E.D. Tex. Jan. 5, 2016) (Payne, J.) (“‘processor’ connotes structure”); *see also id.*, at \*18–\*21. Also, the Federal Circuit has reinforced this analysis, finding that a “processing” term connoted structure:

As used in the claims of the ’591 patent, the term “digital processing unit” clearly serves as a stand-in for a “general purpose computer” or a “central processing unit,” each of which would be understood as a reference to structure in this case, not simply any device that can perform a particular function.

*Samsung Elecs. Am., Inc. v. Prisia Eng’g Corp.*, 948 F.3d 1342, 1354 (Fed. Cir. 2020). The opinions of Defendant’s expert to the contrary are unpersuasive. (*See* Dkt. #62, Ex. 1, Sept. 13, 2024 Weissman Decl. ¶¶ 56–61).

Defendant’s reliance on *Advanced Ground Information Systems* is similarly unpersuasive. *See Advanced Ground Info. Sys., Inc. v. Life360, Inc.*, 830 F.3d 1341, 1347–48 (Fed. Cir. 2016) (“The term ‘symbol generator’ invokes the application of § 112, ¶ 6 because it fails to describe a sufficient structure and otherwise recites abstract elements ‘for’ causing actions . . . or elements ‘that can’ perform functions.”). As discussed above, “processor” and “memory” connote structure in the present case.

Defendant also cites the Federal Circuit’s non-precedential opinion in *WSOU Investments LLC v. Google LLC*, Nos. 2022-1063, 2022-1065, 2023 WL 6889033, at \*3 (Fed. Cir. Oct. 19, 2023). As to the term there at issue—“said processor configured to provide a pre-emptive user output when the sub-set of pixels approaches an edge of the set of available pixels”—the Federal Circuit applied § 112, ¶ 6, noting that “[i]n this case, as the district court correctly noted, the specification treats the word ‘processor’ so broadly as to generically be any structure that manipulates data.” *Id.*, at \*4. No such showing has been made in the present case. On the contrary,

the specification uses the terms “processor” and “memory” to refer to classes of structures. The non-precedential *VDPP* case cited in Plaintiff’s reply brief, as well as the precedential *Dyfan* decision discussed in *VDPP*, are of further persuasive value in this regard. *VDPP, LLC v. Vizio, Inc.*, No. 2021-2040, 2022 WL 885771, at \*3–\*4 (Fed. Cir. Mar. 25, 2022) (reversing means-plus-function treatment of “a processor adapted to”) (discussing *Dyfan, LLC v. Target Corp.*, 28 F.4th 1360 (Fed. Cir. 2022) (finding that “code” and “application” limitations were not means-plus-function)).

Because “processor” and “memory” connote structure, Defendant’s reliance on authorities regarding functional configuration are inapplicable. *See MTD Prods. Inc. v. Iancu*, 933 F.3d 1336, 1343 (Fed. Cir. 2019) (“[T]he claim language reciting what the mechanical control assembly is ‘configured to’ do is functional.”); *see also Rain Computing, Inc. v. Samsung Elecs. Am., Inc.*, 989 F.3d 1002, 1006 (Fed. Cir. 2021) (“[T]he purely functional claim language reciting what the ‘user identification module’ is configured to do provides no structure.”). Also, because § 112, ¶ 6 does not apply, Defendant’s cited authority regarding general-purpose processors as corresponding structure (and the algorithm requirement associated therewith) is inapplicable. *See In re Katz Interactive Call Processing Pat. Litig.*, 639 F.3d 1303, 1316 (Fed. Cir. 2011).

At the December 2, 2024 hearing, Defendant urged that the claim language merely recites results, analogizing to Sidney Harris’s cartoon depicting a mathematical proof that includes “Then a Miracle Occurs . . .,” but while Defendant’s argument perhaps might pertain to a challenge regarding enablement or written description, Defendant’s argument does not compel any finding of means-plus-function or indefiniteness because, as discussed above, “processor and memory” connote structure.

The Court therefore hereby expressly rejects Defendant’s argument that this is a means-plus-function term governed by 35 U.S.C. § 112, ¶ 6. Defendant presents no alternative proposed construction, and no further construction is necessary.

The Court accordingly hereby construes this term to have its **plain meaning**.

**2. “mapping” / “map” / “mapped”**

<b>“mapping” “map” “mapped” (Claims 1, 17)</b>	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
Plain and ordinary meaning.	“[processing] / [process] / [processed] key / value pairs to generate intermediate key / value pairs”

(Dkt. #51, Ex. A, at p. 2; *id.*, Ex. B, at p. 3; Dkt. #53, at p. 10; Dkt. #62, at p. 17; Dkt. #69, Ex. A, at p. 3).

**a. The Parties’ Positions**

Plaintiff argues that “the claims explain what the mapping functions do and the mapping that they perform,” and “[t]he claims are also consistent with the specification, which is replete with examples demonstrating what mapping functions are and what they do.” (Dkt. #53, at p. 10) (citations omitted). Plaintiff also argues that Defendant’s proposed constructions would render other claim language superfluous. (*Id.*, at p. 11).

Defendant responds that “the applicants acted as their own lexicographer,” and “the claims and specification confirm that special meaning.” (Dkt. #62, at p. 18). Defendant also argues that “R2’s attempt to redefine map under the guise of plain meaning is not only inconsistent with the

claims and specification, but it will create jury confusion about the meaning of this technical term.” (*Id.*, at p. 20) (citation omitted).

Plaintiff replies that “[m]erely explaining what a map function ‘basically’ does in the background section of the patent falls well short of th[e] [lexicography] standard.” (Dkt. #66, at p. 6).

### **b. Analysis**

Claim 1 of the ’610 Patent, for example, recites (emphasis added):

1. A method of processing data of a data set over a distributed system, wherein the data set comprises a plurality of data groups, the method comprising:

partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of *mapping* functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such *map* function’s corresponding data partition *to form corresponding intermediate data* for that data group and identifiable to that data group, wherein the data of a first data group has a different schema than the data of a second data group and the data of the first data group is *mapped* differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different schema and corresponding different intermediate data have a key in common; and

reducing the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group, so as to result in a merging of the corresponding different intermediate data based on the key in common,

wherein the *mapping* and reducing operations are performed by a distributed system.

The Background section of the specification states:

MapReduce is a programming methodology to perform parallel computations over distributed (typically, very large) data sets. Some theory regarding the MapReduce programming methodology is described in “MapReduce: Simplified Data Processing on Large Clusters,” by Jeffrey Dean and Sanjay Ghemawat, appearing in OSDI’04: Sixth Symposium on Operating System Design and Implementation, San Francisco, Calif., December, 2004 (hereafter, “Dean and Ghemawat”). A similar, but not identical, presentation is also provided in HTML form at the following URL: <http://labs.google.com/papers/mapreduce-osdi04-slides/index.html> (hereafter, “Dean and Ghemawat HTML”).

Basically, a “map” function maps key-value pairs to new (intermediate) key-value pairs. A “reduce” function represents all mapped (intermediate) key-value pairs sharing the same key to a single key-value pair or a list of values. The “map” and “reduce” functions are typically user-provided. The map function iterates over a list of independent elements, performing an operation on each element as specified by the map function. *The map function generates intermediate results.* The reduce operation takes these intermediate results via a single iterator and combines elements as specified by the reduce function.

’610 Patent at 1:6–27 (emphasis added); *see id.* at 2:21–24 (“The programmer specifies a map function that processes input key/value pairs and produces a set of intermediate pairs 106(1) through 106(7).”).

A term being “set off by quotation marks” in the specification is “often a strong indication that what follows is a definition.” *Sinorgchem Co., Shandong v. Int’l Trade Comm’n*, 511 F.3d 1132, 1136 (Fed. Cir. 2007) (citation omitted).

Also of note, the “Dean and Ghemawat” reference cited here by the Background refers to “map” similarly, stating in its Abstract that “[u]sers specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs . . . .” (Dkt. #62, Ex. 2, Jeffery Dean & Sanjay Ghemawat, *MapReduce: Simplified Data Processing on Large Clusters* (2004), at Abstract) (emphasis omitted); *see Powell v. Home Depot U.S.A., Inc.*, 663 F.3d 1221, 1231 (Fed. Cir. 2011) (“prior art cited in a patent or cited in the prosecution history of the patent constitutes intrinsic evidence”).

Defendant also emphasizes authority holding that the proper construction of a term is the definition provided in the specification, despite a risk of redundancy, where “intrinsic evidence makes it clear that the ‘redundant’ construction is correct.” *VLSI Tech. v. Intel Corp.*, 53 F.4th 646, 653 (Fed. Cir. 2022).

Yet, disclosure that a “‘map’ function” is something that “maps” is not a lexicography as to the term “map.” In the context of this disclosure, using the word “map” to purportedly define a



“‘map’ function” “does not provide reasonable clarity, deliberateness, and precision sufficient to narrow the definition of the claim term in the manner urged.” *Abbott Labs v Syntron Bioresearch*, 334 F.3d 1343, 1355 (Fed. Cir. 2003).

Also, surrounding claim language already provides ample context for understanding the plain meaning of these disputed terms as used in the claims at issue. Indeed, Defendant’s proposed construction is also already largely covered by other claim language, such as the recital of “form[ing] corresponding intermediate data,” and would tend to confuse rather than clarify the scope of the claims. *See, e.g., Apple, Inc. v. Ameranth, Inc.*, 842 F.3d 1229, 1237 (Fed. Cir. 2016) (“Construing a claim term to include features of that term already recited in the claims would make those expressly recited features redundant.”).

Finally, as to Defendant’s concern regarding Plaintiff’s interpretation of the plain meaning of these terms as being to “filter and/or sort data” (Dkt. #53, at p. 1) (“MapReduce methods consist of a map procedure to filter and/or sort data”), Plaintiff does not propose this as a construction, and any dispute regarding Plaintiff’s application of the claim language in an infringement analysis would appear to pertain to factual disputes regarding infringement rather than any legal question for claim construction. *See Acumed LLC v. Stryker Corp.*, 483 F.3d 800, 806 (Fed. Cir. 2007) (“[t]he resolution of some line-drawing problems . . . is properly left to the trier of fact”) (citing *PPG Indus. v. Guardian Indus. Corp.*, 156 F.3d 1351, 1355 (Fed. Cir. 1998)).

The Court therefore hereby construes “mapping,” “map,” and “mapped” to have their **plain meaning**.

### 3. “reducing” / “reduce”

<p style="text-align: center;"><b>“reducing”</b> <b>“reduce”</b> (Claims 1, 17)</p>	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
Plain and ordinary meaning.	“[merging] / [merge] all intermediate data values sharing the same key into a single key-value pair or a list of values associated with the key” <sup>1</sup>

(Dkt. #51, Ex. A, at p. 3; *id.*, Ex. B, at p. 4; Dkt. #53, at p. 11; Dkt. #57, at p. 21; Dkt. #69, Ex. A, at p. 4).

#### a. The Parties’ Positions

Plaintiff argues that “the claims explain what the reducing does (e.g., converts intermediate data to at least one output data group), how it is accomplished (e.g., processing in a data-group specific manner), and what the practical result is (e.g., merging of the intermediate data based on the common key),” and “[t]here is no reason to construe terms that are exhaustively defined in the claims themselves.” (Dkt. #53, at p. 12). “Furthermore,” Plaintiff argues, “reducing is thoroughly discussed in the specification, and no POSA would be confused as to its meaning in light thereof.” (*Id.*) (citations omitted).

Defendant responds that “the applicants acted as their own lexicographer by defining the term ‘reduce’ . . . .” (Dkt. #62, at p. 21). Defendant also argues: “[T]he examples R2 identifies—such as ‘counting’ and ‘frequency determination’—are examples of how MapReduce may be applied. These are not examples of what ‘reduce’ is.” (*Id.*, at p. 22) (citation omitted).

---

<sup>1</sup> In its Responsive Claim Construction Brief, Defendant amended its proposal so as to replace “combining” and “combine” with “merging” and “merge,” respectively.

Plaintiff replies that “[t]he background explanation regarding typical MapReduce does not constitute a clear intent by the patentee to define the terms, which is demonstrated by the fact that Databricks’ proposal is not verbatim from the specification and is instead a mix-and-match of whatever disclosure Databricks saw fit to cram into its construction.” (Dkt. #66, at pp. 7–8). At the December 2, 2024 hearing, Plaintiff also urged that other claim language already addresses what is set forth in the purported lexicography, and Plaintiff also noted that no lexicography was proposed by the parties in prior litigation involving this patent.

### **b. Analysis**

Claim 1 of the ’610 Patent, for example, recites (emphasis added):

1. A method of processing data of a data set over a distributed system, wherein the data set comprises a plurality of data groups, the method comprising:

partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of mapping functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such map function’s corresponding data partition to form corresponding intermediate data for that data group and identifiable to that data group, wherein the data of a first data group has a different schema than the data of a second data group and the data of the first data group is mapped differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different schema and corresponding different intermediate data have a key in common; and

*reducing the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group, so as to result in a merging of the corresponding different intermediate data based on the key in common,*

wherein the mapping and *reducing* operations are performed by a distributed system.

The Background section of the specification states:

MapReduce is a programming methodology to perform parallel computations over distributed (typically, very large) data sets. Some theory regarding the MapReduce programming methodology is described in “MapReduce: Simplified Data Processing on Large Clusters,” by Jeffrey Dean and Sanjay Ghemawat, appearing in OSDI’04: Sixth Symposium on Operating System Design and Implementation, San Francisco, Calif., December, 2004 (hereafter, “Dean and Ghemawat”). A

similar, but not identical, presentation is also provided in HTML form at the following URL: <http://labs.google.com/papers/mapreduce-osdi04-slides/index.html> (hereafter, “Dean and Ghemawat HTML”).

Basically, a “map” function maps key-value pairs to new (intermediate) key-value pairs. *A “reduce” function represents all mapped (intermediate) key-value pairs sharing the same key to a single key-value pair or a list of values.* The “map” and “reduce” functions are typically user-provided. The map function iterates over a list of independent elements, performing an operation on each element as specified by the map function. The map function generates intermediate results. *The reduce operation takes these intermediate results via a single iterator and combines elements as specified by the reduce function.*

’610 Patent at 1:6–27 (emphasis added). Reduce functions are also discussed in the Detailed Description section of the specification. *See, e.g., id.* at 2:49–51 (“The reduce function combines all intermediate values for a particular key and produces a set of *merged output values* for the key.”) (emphasis added); *id.* at 5:40–6:17.

Unlike for the above-discussed “map” terms, the above-reproduced Background does not use “reduce” to describe “reduce” but rather sets forth a definition of “reduce” as “represents all mapped (intermediate) key-value pairs sharing the same key to a single key-value pair or a list of values.” *Id.* at 1:18–20.

Particularly in light of the patentee’s usage of quotation marks around the word “reduce,” this amounts to a lexicography for the term “reduce.” *Sinorgchem*, 511 F.3d at 1136; *see Phillips*, 415 F.3d at 1316 (“our cases recognize that the specification may reveal a special definition given to a claim term by the patentee that differs from the meaning it would otherwise possess,” and “[i]n such cases, the inventor’s lexicography governs”).

Also of note, the “Dean and Ghemawat” reference cited here by the Background refers to “reduce” similarly, stating in its Abstract that a “reduce function” “merges all intermediate values associated with the same intermediate key.” (Dkt. #62, Ex. 2, Jeffery Dean & Sanjay Ghemawat,

*MapReduce: Simplified Data Processing on Large Clusters* (2004), at Abstract) (emphasis omitted).

To whatever extent applying the patentee’s lexicography results in some amount of redundancy in the claims, such redundancy is permissible and does not override the patentee’s lexicography. *See VLSI*, 53 F.4th at 653 (“intrinsic evidence makes it clear that the ‘redundant’ construction is correct”).

During the December 2, 2024 hearing, the Court proposed construing “reducing” to mean “representing all mapped (intermediate) key-value pairs sharing the same key to a single key-value pair or a list of values,” and Defendant was amenable to that construction. Later during the hearing, the Court also suggested replacing “to a single key-value pair or a list of values” with “as a single key-value pair or a list of values” to aid clarity, and, although Defendant was not prepared to agree to that change, Defendant had no objection thereto. Having considered this again post-hearing, the Court believes that this minor clarification of the patentee’s lexicography will assist the finder of fact.

The Court therefore hereby construes these terms as set forth in the following chart:

<u>Term</u>	<u>Construction</u>
“reducing”	<b>“representing all mapped (intermediate) key-value pairs sharing the same key as a single key-value pair or a list of values”</b>
“reduce”	<b>“represent all mapped (intermediate) key-value pairs sharing the same key as a single key-value pair or a list of values”</b>

**4. “providing each data partition to a selected one of a plurality of mapping functions”**

<b>“providing each data partition to a selected one of a plurality of mapping functions”</b> (Claims 1, 17)	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
Plain and ordinary meaning.	“providing each data partition to one of a plurality of different mapping functions where each mapping function is selected for a partition based on the data group the partition originated from” <sup>2</sup>

(Dkt. #51, Ex. A, at p. 4; *id.*, Ex. B, at p. 5; Dkt. #53, at p. 13; Dkt. #62, at p. 22; Dkt. #69, Ex. A, at pp. 5–6).

**a. The Parties’ Positions**

Plaintiff argues that “the language [from Inter Partes Review (‘IPR’) proceedings] that Databricks relies upon is not disclaimer,” and “[m]oreover, Databricks’ proposal is confusing and grammatically inaccurate.” (Dkt. #53, at p. 13). Plaintiff concludes: “The plain language of the claims simply necessitates that each partition be provided ‘to a selected one of a plurality of mapping functions,’ as opposed to multiple partitions going to a single mapping function. This is made clear throughout the specification. *See, e.g.*, ’610 pat. at Figs. 4, 5. No construction is necessary.” (*Id.*, at p. 17).

Defendant responds that Defendant’s proposed construction “is required by R2’s clear prosecution history disclaimer during ’610 patent IPR proceedings.” (Dkt. #62, at p. 22; *see id.*, at pp. 22–26).

---

<sup>2</sup> In its Responsive Claim Construction Brief, Defendant amended its proposal so as to replace “the mapping function” with “each mapping function.” (Dkt. #62, at p. 22).

Plaintiff replies by emphasizing that Defendant changed its proposed construction in its responsive brief, arguing that “Databricks has now offered two different interpretations of the same statement, which should, by itself, put disclaimer to rest.” (Dkt. #66, at p. 8). Further, Plaintiff argues that “[w]hile R2 argued that the prior art does not teach delineating data based on data groups, *nowhere* did it argue that the prior art failed to teach that a mapping function is ‘selected for a partition based on the data group the partition originated from.’” (*Id.*, at pp. 8–9).

At the December 2, 2024 hearing, Plaintiff reiterated its arguments that the IPR statement cited by Defendant appeared in general summary comments and was not tethered to any claim limitation. Plaintiff also argued that the present disputed term was specifically addressed in this IPR document without any statement that would support Defendants’ proposed construction. Defendant responded that the statement at issue was explicitly directed to the scope of the claims, that the patentee relied on this interpretation when distinguishing prior art, and that the PTAB subsequently used the same interpretation.

#### **b. Analysis**

“Claims may not be construed one way in order to obtain their allowance and in a different way against accused infringers.” *Southwall Techs., Inc. v. Cardinal IG Co.*, 54 F.3d 1570, 1576 (Fed. Cir. 1995). “Because an IPR proceeding involves reexamination of an earlier administrative grant of a patent, it follows that statements made by a patent owner during an IPR proceeding can be considered during claim construction and relied upon to support a finding of prosecution disclaimer.” *Aylus Networks, Inc. v. Apple Inc.*, 856 F.3d 1353, 1360 (Fed. Cir. 2017). Statements are considered “in the context of the entire prosecution history.” *Tech. Props. Ltd. LLC v. Huawei Techs. Co., Ltd.*, 849 F.3d 1349, 1357–58 (Fed. Cir. 2017) (citation omitted). “[W]hen a

prosecution argument is subject to more than one reasonable interpretation, it cannot rise to the level of a clear and unmistakable disclaimer.” *Aylus*, 856 F.3d at 1363 (citation omitted).

During IPR proceedings, Plaintiff stated as follows in relation to the “Pike” reference and the “Chowdhuri” reference:

VIII. THE PETITION DOES NOT DEMONSTRATE THAT THE CHALLENGED CLAIMS ARE UNPATENTABLE.

To justify the institution of an inter partes review, a petitioner must establish that there is a “reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition.” 35 U.S.C. § 314(a). Petitioner has not met its burden of establishing a reasonable likelihood of success at least because the Petition fails to show that each limitation of the independent claims is disclosed or taught by the asserted prior art references of each ground. Consequently, the Board should deny the Petition and decline to institute the inter partes review.

As discussed above, *the ’610 patent claims* a stem-to-stern augmentation to MapReduce, requiring specialized data structures (“data groups”) at the front-end configured to be merged together via selective dissemination amongst and processing by a collection of mapping and reducing functions within the enhanced MapReduce architecture. This in turn *requires the mapping functions to be specialized so that they can be “selected” for a partition based on the “data group” the partition originated from and map the partitions “differently” based on “data group”* to form “corresponding intermediate data.” It also requires a reducing phase configured to distinguish the “intermediate data” based on the parent “data groups” to effectuate specialized processing on a per-“data group” basis. These enhancements are reflected as requirements of each independent claim. *See* Ex. 1026. Claims 17 and 40 mirror the requirements of Claims 1 and 33, respectively. *See id.*

These features are not disclosed by Pike or Chowdhuri. \* \* \*

(Dkt. #53, Ex. 4, May 22, 2024 Patent Owner’s Preliminary Response, at pp. 35–36) (emphasis added). Plaintiff further stated:

[A]n understanding of Chowdhuri’s methodology emphasizes why Chowdhuri does not disclose intermediate data that is identifiable to a data group—Chowdhuri does not need a “mechanism” to identify which parent data group intermediate data came from because Chowdhuri accomplishes data consolidation via the same function, i.e., the hash join. *This is different from the ’610 patent, which employs “data groups” that specialized map and reduce functions can use to delineate between data (including “intermediate data”) that originated from one group as*



*compared to another to properly distribute and merge the data across the parallelized architecture.* \* \* \* Chowdhuri does not disclose “intermediate data” that is “identifiable” to a parent “data group” because Chowdhuri has no such “intermediate data” and needs no such identifier.

*Id.* at 65–66 (emphasis added); *see id.* at 57–58 (“There is no disclosure in Chowdhuri of any sort of ‘identifier (table name or otherwise) being, e.g., passed to specific functions to guide group-specific processing as in the ’610 patent.’”) (citations omitted).

Plaintiff argues that these statements do not contain any clear disclaimer, arguing that “R2 was simply explaining that the ‘enhanced MapReduce architecture . . . *requires the mapping functions to be specialized,*’ not that the mapping functions are required to be selected for partitions based on data group.” (Dkt. #53, at p. 14). Plaintiff for example cites a finding of no disclaimer in *VARTA Microbattery GmbH v. Audio P’ship LLC*, No. 2:21-CV-400-JRG-RSP, 2023 WL 5103113, at \*4 (E.D. Tex. Aug. 9, 2023).

As Defendant points out, however, the IPR statements here at issue do not contain any of the qualifiers that were present in *VARTA*, such as “possible,” “example,” and “preferred.” *See id.*

Moreover, the Patent Trial and Appeal Board (“PTAB”) appears to have relied on Plaintiff’s assertions, noting in its June 11, 2024 Decision Denying Institution of Inter Partes Review that “[t]he ‘data group’ identifiers in the ’610 patent identify input data sets with file names for different schemas or data types, which allows a programmer to program and execute MapReduce functions that correspond to a particular ‘data group’ when joining two different types of input data based on a common key.” (Dkt. #53, Ex. 5, at 30). Alternatively, even if the PTAB’s Decision is interpreted such that the PTAB did not rely on any relevant statement by Plaintiff, Plaintiff’s statements can support a finding of disclaimer regardless of whether the PTAB relied upon them. *Cf. Seachange Int’l, Inc. v. C–COR Inc.*, 413 F.3d 1361, 1374 (Fed. Cir. 2005) (“An

applicant's argument made during prosecution may lead to a disavowal of claim scope even if the Examiner did not rely on the argument.").

At the December 2, 2024 hearing, Plaintiff reiterated the arguments set forth in its briefing that the statements at issue appeared in an introductory section of this IPR document, that the statements were not directed to any particular claim limitation, and that the disputed term here at issue was specifically addressed elsewhere in this IPR document without any discussion of each mapping function being selected for a partition based on the data group the partition originated from.

Plaintiff cites no authority for its suggestion that the location of the statements within the document can undercut a finding of disclaimer. Indeed, Plaintiff's reliance on the location of this statement as being "part of a general summary of certain enhancements that Databricks' prior art did not disclose" (Dkt. #53, at p. 15) merely reinforces that the statements at issue limit the claimed invention as a whole. Indeed, these statements expressly refer to what "the '610 patent claims," do so in the context of distinguishing specific prior art (Pike and Chowdhuri), and expressly reiterate that "[t]hese enhancements are reflected as requirements of each independent claim." Moreover, Plaintiff shows no requirement that a statement during prosecution or IPR must be specifically tethered to a particular limitation within a claim (rather than to a claim as a whole or to the claims of the patent as a whole).

Plaintiff also argues that the statement was "exemplary and intended only to provide helpful explanation." (Dkt. #53, at p. 14). Plaintiff's argument is contrary to the "basic principle of claim interpretation" that "prosecution disclaimer promotes the *public notice* function of the intrinsic evidence and protects the *public's reliance* on definitive statements made during prosecution." *Omega Eng'g*, 334 F.3d at 1324 (citation omitted; emphasis added). The inquiry is

whether Plaintiff set forth to the public a clear and unmistakable statement regarding the claim scope. As discussed above, Plaintiff did so. Plaintiff's purported intent does not undercut the finding of disclaimer.

Finally, Plaintiff cites the principle that “when a prosecution argument is subject to more than one reasonable interpretation, it cannot rise to the level of a clear and unmistakable disclaimer.” *Aylus*, 856 F.3d at 1363. Like in *Aylus*, however, Plaintiff's alternative interpretation of its IPR statements, as requiring nothing more than the mapping functions being specialized, is “not reasonable” and therefore does not avoid a finding of disclaimer. *Id.* Plaintiff's statements in IPR rise to the level of a “clear and unmistakable” disclaimer of claim scope in the manner discussed above. *Id.* at 1359 (citation omitted).

Finally, whereas Plaintiff criticizes Defendant's proposed construction as confusing and as purportedly lacking support in the cited IPR statements, Defendant's proposal is reasonably clear and gives proper effect to Plaintiff's above-discussed IPR statements.

The Court therefore hereby construes “providing each data partition to a selected one of a plurality of mapping functions” to mean “**providing each data partition to one of a plurality of different mapping functions where each mapping function is selected for a partition based on the data group the partition originated from.**”

**5. “[processing] / [process] the intermediate data for each data group in a manner that is defined to correspond to that data group”**

<b>“[processing] / [process] the intermediate data for each data group in a manner that is defined to correspond to that data group”</b> (Claims 1, 5, 17, 21)	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
Plain and ordinary meaning	“[processing] / [process] the intermediate data for each data group in a manner that is defined to correspond to the data group from which the intermediate data originated”

(Dkt. #51, Ex. A, at p. 4; *id.*, Ex. B, at pp. 5–6; Dkt. #53, at p. 17; Dkt. #69, Ex. A, at p. 6).

**a. The Parties’ Positions**

Plaintiff submits that the Court previously construed these terms to have their plain meaning, and Plaintiff argues that “Databricks’ construction is unnecessary and simply uses more words to describe what the limitation already concisely says.” (Dkt. #53, at p. 17).

Defendant responds that, in IPR proceedings, Plaintiff interpreted “that data group” in the same manner that Defendant proposes here. (Dkt. #62, at pp. 26–27). Defendant also submits that, in *Walmart*, “neither party sought to clarify the meaning of ‘that data group,’” and “the Court did not have before it R2’s May 22, 2024 IPR statements expressly clarifying the meaning of ‘that data group’ consistent with Databrick’s proposed construction.” (*Id.*, at p. 28) (citation omitted). Defendant urges that “[b]ecause the claims refer to at least three different data groups and ‘that data group’ is ambiguous in the context of the claims, construing ‘that data group’ consistent with the parties’ agreed clarification will clarify the scope of the claims for the jury.” (*Id.*)

Plaintiff replies: “There is no reason to add constructions when the plain claim language will do. Replacing ‘that data group’ with ‘the data group from which the intermediate data

originated’ adds nothing but extra words. The phrases mean the same thing. No construction of this phrase is necessary.” (Dkt. #66, at p. 10).

**b. Analysis**

Claim 1 of the ’610 Patent, for example, recites (emphasis added):

1. A method of processing data of a data set over a distributed system, wherein the data set comprises a plurality of data groups, the method comprising:

partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of mapping functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such map function’s corresponding data partition to form corresponding intermediate data for that data group and identifiable to that data group, *wherein the data of a first data group has a different schema than the data of a second data group* and the data of the first data group is mapped differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different schema and corresponding different intermediate data have a key in common; and

*reducing the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group*, so as to result in a merging of the corresponding different intermediate data based on the key in common,

wherein the mapping and reducing operations are performed by a distributed system.

In *Walmart*, the Court construed “including processing the intermediate data for each data group in a manner that is defined to correspond to that data group, so as to result in a merging of the corresponding different intermediate data based on the key in common” in Claim 1 of the ’610 Patent to have its plain meaning. *Walmart* at 27–29. The parties in *Walmart* did not raise any issue regarding this recital of “that data group.” *See id.*

In the present case, Defendant proposes a construction to explain that this recital of “that data group” refers to the data group from which the intermediate data originated.

Defendant’s proposal is consistent with the claim language and will assist the finder of fact in understanding this limitation. Defendant’s proposal is also consistent with Plaintiff’s apparent understanding in IPR proceedings, in which Plaintiff stated:

[The] Pike [reference] does not disclose “data groups” that have a “key in common” with each other and the “intermediate data,” so Pike cannot teach “merging” based on such a common key. Further, Pike does not disclose that processing “intermediate data” in the reduce phase occurs *“in a manner that is defined to correspond” to the “data groups” from which the “intermediate data” originated.*

(Dkt. #53, Ex. 4, May 22, 2024 Patent Owner’s Preliminary Response, at p. 55) (emphasis added).

At the December 2, 2024 hearing, Plaintiff argued that Defendant’s proposed construction is unnecessary but expressed no substantive opposition to the construction.

The Court therefore hereby construes “[processing] / [process] the intermediate data for each data group in a manner that is defined to correspond to that data group” to mean “[**processing**] / [**process**] the intermediate data for each data group in a manner that is defined to correspond to the data group from which the intermediate data originated.”

**6. “schema”**

<b>“schema”</b> (Claims 1, 17)	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
Plain and ordinary meaning.	“a set of attributes (such as DeptID, LastName, DeptName) and their properties (such as their data types: integer DeptID, string LastName, string DeptName)”

(Dkt. #51, Ex. A, at p. 5; *id.*, Ex. B, at p. 6; Dkt. #53, at p. 18; Dkt. #69, Ex. A, at p. 8; Dkt. #69, Ex. A, at p. 8).

**a. The Parties' Positions**

Plaintiff argues that “[s]chema” is a readily-understandable term that all POSAs know, and the specification is rife with disclosure that would enable a POSA, and the jury, to understand the term’s meaning within the patent’s context,” and “the specification even provides explicit examples of data of different schema.” (Dkt. #53, at p. 18) (citations omitted). Plaintiff also argues that “Databricks’ proposed construction is an obvious attempt to narrow the meaning of the term to a single embodiment.” (*Id.*)

Defendant responds that “the concept of different schemas goes to the heart of the applicants’ purported invention,” and “[t]o explain how the ’610 patent’s improved MapReduce is different, the ’610 patent defines ‘schema’ in the context of the tables of Figure 3.” (Dkt. #62, at p. 29). Defendant also argues that its proposal is supported by an extrinsic technical dictionary. (*Id.*)

Plaintiff replies that “R2 has offered to agree to construe the term as ‘a set of attributes,’ but Databricks is keen to make the term seem narrower than it is by keeping the phrase ‘(such as DeptID, LastName, DeptName).’” (Dkt. #66, at p. 10). Plaintiff urges that “[e]very POSA knows what ‘schema’ means, and supplanting a single word with a convoluted two-dozen-word explanation does not help the jury, particularly when the context of the patent is more than sufficient to aid in understanding the term’s plain and ordinary meaning.” (*Id.*).

At the December 2, 2024 hearing, Defendant reiterated its arguments that having properties associated with attributes is inherent in a “schema,” that this is the only disclosed embodiment, and that the specification uses this embodiment to explain what a “schema” is. Defendant also argued that Plaintiff is attempting to interpret “schema” so broadly that it encompasses unstructured data (data that, in Defendant’s view, has no schema at all).

**b. Analysis**

Claim 1 of the '610 Patent, for example, recites (emphasis added):

1. A method of processing data of a data set over a distributed system, wherein the data set comprises a plurality of data groups, the method comprising:

partitioning the data of each one of the data groups into a plurality of data partitions that each have a plurality of key-value pairs and providing each data partition to a selected one of a plurality of mapping functions that are each user-configurable to independently output a plurality of lists of values for each of a set of keys found in such map function's corresponding data partition to form corresponding intermediate data for that data group and identifiable to that data group, wherein the data of a first data group has a different *schema* than the data of a second data group and the data of the first data group is mapped differently than the data of the second data group so that different lists of values are output for the corresponding different intermediate data, wherein the different *schema* and corresponding different intermediate data have a key in common; and

reducing the intermediate data for the data groups to at least one output data group, including processing the intermediate data for each data group in a manner that is defined to correspond to that data group, so as to result in a merging of the corresponding different intermediate data based on the key in common,

wherein the mapping and reducing operations are performed by a distributed system.

Defendant argues that the specification defines the term “schema” in the following disclosure:

FIG. 4 illustrates an example of an improved MapReduce architecture in accordance with an aspect, and relative to the FIG. 3 example relational tables. It is noted that *the schema of each data set, such as the FIG. 3 relational tables, includes a set of attributes (such as DeptID, LastName, DeptName) and their properties (such as their data types: integer DeptID, string LastName, string DeptName)*. As discussed above with respect to the conventional MapReduce architecture, input, intermediate and output data sets may each be characterized by their own schema, and each schema operates according to “key/value” pairs. The attributes in the key/value pairs may be distinct or overlapping; moreover, keys within a data set may be distinct or identical.

'610 Patent at 3:35–47 (emphasis added); *see id.* at 3:48–64.

This disclosure, regarding what a schema “includes,” appears as part of a discussion of embodiments. Defendant argues that “the patent uses Figure 3 to define what a ‘schema’ is and illustrate how the improved MapReduce is purportedly different from conventional MapReduce



based on this definition” (Dkt. #62, at p. 29 n.22), and Defendant submits authority that a definition can appear as part of a disclosure of a particular embodiment. *See Kyocera Senco Indus. Tools Inc. v. Int’l Trade Comm’n*, 22 F.4th 1369, 1376–79 (Fed. Cir. 2022). This discussion of “schema,” however, does not state what the term is “defined as” or what the term “is” and does not use quotation marks around the term. *See, e.g., Abbott*, 334 F.3d at 1354. Defendant does not persuasively demonstrate “reasonable clarity, deliberateness, and precision” so as to rise to the level of a lexicography. *Id.*

Further, although much of Defendant’s proposed construction is non-limiting examples, any benefit from including those examples in the construction is outweighed by the risk that the finder of fact might perceive the examples as limiting.

Finally, to the extent Defendant is arguing that a schema necessarily has “properties,” if this is indeed a technical necessity then it is inherent and therefore need not be included in the Court’s construction. Defendant’s reliance on an extrinsic technical dictionary definition is likewise unpersuasive. (*See* Dkt. #62, Ex. 7, *Microsoft Computer Dictionary* 465 (5th ed. 2002)) (“A description of a database to a database management system (DBMS) in the language provided by the DBMS. A schema defines aspects of the database, such as attributes (fields) and domains and parameters of the attributes.”).

The Court therefore hereby expressly rejects Defendant’s proposed construction, and no further construction is necessary. *See O2 Micro*, 521 F.3d at 1362 (“[D]istrict courts are not (and should not be) required to construe every limitation present in a patent’s asserted claims.”); *see also Finjan, Inc. v. Secure Computing Corp.*, 626 F.3d 1197, 1207 (Fed. Cir. 2010) (“Unlike *O2 Micro*, where the court failed to resolve the parties’ quarrel, the district court rejected Defendants’ construction.”); *Bayer Healthcare LLC v. Baxalta Inc.*, 989 F.3d 964, 977–79 (Fed. Cir. 2021).

The Court accordingly hereby construes “**schema**” to have its **plain meaning**.

**7. “the different schema and corresponding different intermediate data have a key in common”**

<b>“the different schema and corresponding different intermediate data have a key in common”</b> (Claims 1, 17)	
<b>Plaintiff’s Proposed Construction</b>	<b>Defendant’s Proposed Construction</b>
R2 Solutions contends that this phrase needs no construction and should be afforded its plain and ordinary meaning.	“the different data group schemas have a key in common with the corresponding different intermediate data”

(Dkt. #51, Ex. A, at p. 6; *id.*, Ex. B, at p. 7; Dkt. #53, at p. 19; Dkt. #69, Ex. A, at p. 8).

Plaintiff submits that “[t]he parties agree that the phrase ‘the different schema and corresponding different intermediate data have a key in common’ should be construed to mean ‘the different data group schemas have a key in common with the corresponding different intermediate data.’” (Dkt. #53, at p. 19).

Defendant responds that “R2 agrees to Databricks’ proposed construction, which resolves the dispute related to this term,” and “[t]hus, the Court should adopt the parties’ agreed construction.” (Dkt. #62, at p. 30) (citing Dkt. #53, at p. 19).

The Court therefore hereby construes “the different schema and corresponding different intermediate data have a key in common” to mean **“the different data group schemas have a key in common with the corresponding different intermediate data.”**

## CONCLUSION

The Court adopts the constructions set forth in this opinion for the disputed terms of the patents-in-suit. The parties are ordered that they may not refer, directly or indirectly, to each other’s claim construction positions in the presence of the jury. Likewise, the parties are ordered

to refrain from mentioning any portion of this opinion, other than the actual definitions adopted by the Court, in the presence of the jury. Any reference to claim construction proceedings is limited to informing the jury of the definitions adopted by the Court.

**IT IS SO ORDERED.**

**SIGNED this 10th day of December, 2024.**

A handwritten signature in black ink, reading "Amos Mazzant". The signature is written in a cursive, flowing style. Below the signature is a horizontal line.

AMOS L. MAZZANT  
UNITED STATES DISTRICT JUDGE